DOCUMENT RESUME

ED 098 541 CS 001 481

AUTHOR Johns, Jerry L.
TITLE Some Comparisons Between the Dolch Basic Sight Vocabulary and the Word List for the 1970's.
PUB DATE [74]
NOTE 12p.; Study conducted at Northern Illinois University

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Readability; Reading; Reading Level; *Reading Research; Vocabulary; *Word Frequency; *Word Lists
IDENTIFIERS Dolch Basic Sight Vocabulary; Word List for the 1970s

ABSTRACT

The purpose of this study was to determine the proportion of words that the Dolch list and the word list for the 1970's accounted for in written materials. The two word lists were compared by token comparison--a method by which words are weighted on the basis of frequency of occurrence within a given set of materials. The written materials used for the comparisons were two large word count studies published within the past decade. In determining the cumulative frequency of words on each list, regularly inflected forms were combined. Calculating the cumulative frequencies of both the word list for the 1970's and the Dolch list in each of the two large word count studies made it possible to determine whether any differences existed and the statistical significance of the differences. The results indicated that the word list for the 1970's accounted for a significantly greater proportion of words than the Dolch list in written materials encountered by both children and adults. However, both lists accounted for over 50 percent of the words used in materials for children and adults. (WR)

SOME COMPARISONS BETWEEN THE DOLCH BASIC SIGHT VOCABULARY AND THE WORD LIST FOR THE 1970'S

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Jerry L. Johns
Northern Illinois University

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Jerry L. Johns
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

ED 098541

There seems to be general agreement among reading authorities that the Dolch (1936) basic sight vocabulary of 220 words has received widespread use among teachers. Recently, however, criticisms (Johnson, 1971b; Otto & Chester, 1972) have characterized the Dolch list with labels such as "passé" and "pseudo-empirical". These criticisms may, at least in part, be justified and there is little wonder that a number of researchers (Harris & Jacobson, 1972; Hillerich, 1974; Johns, 1974; Johnson, 1971a; and Otto & Chester, 1972) have offered word lists to replace or revise Dolch's original sight vocabluary.

Background of Study

Recently, Johns (1974) developed a basic word list for the 1970's which met several criteria. Specifically, the list contained words that occurred frequently in:

1. materials read by children in grades three through nine.
2. materials read by adults.
3. library books read by primary grade children.
4. the spontaneous speaking vocabulary of children in kindergarten and first grade.

In addition, the list contained no nouns and combined regularly inflected forms of a given root word.

CS 001 481

After the word list for the 1970's was compiled, it was assumed that the list would have high utility at all levels of reading development. Not stated, but certainly implied, was the assertion that the word list for the 1970's would be more useful than the Dolch list. One potential measure of "usefulness" is the number and per cent of words in textual materials that can be accounted for by a particular word list. Previous research (Dolch, 1948; Guszak, 1972; Johns, 1971; and Zintz, 1972) has shown that the Dolch list is useful in that it accounts for over 50 per cent of the words in basal readers and other materials. If the word list for the 1970's could account for a significantly greater proportion of words than the Dolch list, there would be an empirical base for claiming that this recently compiled word list is more useful than the Dolch list.

Purpose of Study

The purpose of the present study was to determine the proportion of words that the Dolch list and the word list for the 1970's accounted for in written materials. Specifically, answers to the following two questions were sought:

1. Does the word list for the 1970's account for a greater proportion of words in materials commonly used by children than the Dolch list? Is the difference statistically significant?
2. Does the word list for the 1970's account for a greater proportion of words in materials commonly read by adults than the Dolch list? Is the difference statistically significant?

Procedure for the Comparisons

It was decided that the two word lists would be compared by what is commonly referred to as a token comparison. In a token comparison words are weighted on the basis of their frequency of occurrence within a given set of materials. It was assumed that the written materials used in the study were representative of

those words typically encountered by children and adults. The written materials used for the comparisons were two large word count studies published within the past decade.

The first source utilized was the American Heritage Intermediate (AHI) Corpus published in the Word Frequency Book (Carroll, Davis, & Richman, 1971). The AHI Corpus was compiled from samples of published materials to which U.S. students are exposed in grades three through nine. The materials included textbooks, workbooks, kits, novels, poetry, general nonfiction, encyclopedias, and magazines. The AHI Corpus contains 5,088,721 words drawn in 500-word samples from 1,045 texts. There are 86,741 different words in the Corpus.

The second source was the Kucera-Francis (1967) Corpus. The Corpus was compiled from a wide body of "natural-language" adult published materials ranging from all kinds of newspaper writing to learned journal articles. The Kucera-Francis Corpus contains 1,014,232 words drawn in 500 samples of approximately 2,000 words each. There are 50,406 different words in the Corpus.

In determining the cumulative frequency of words on each list, the investigator combined regularly inflected forms. For this study the term "regularly inflected" included these endings: e, es, ed, er (as comparative, not agent), est, ing, 's (indicating possession or plurality, not contraction), s's, and the dialectal 'in. In general, if the form of the root word was kept intact, the inflected form was included. Examples here would be her-hers, it-its, little-littlest; dialect forms such as know-knowed; and misuses such as best-best's and i for I. Changes in meaning were not included, such as short-shorts, may-May or new-news. Also, two inflected endings (wash-washings) were omitted as well as spelling changes which obliterated the root word (funny-funnier,

ride-riding, sit-sitting). And, finally, archaic forms (the verb endings est, eth), alternate spellings (bye for by), and misspellings were ruled out.

Calculating the cumulative frequencies of the word list for the 1970's and the Dolch list in each of the two large word count studies made it possible to determine whether any differences existed and the statistical significance of the differences.

Results

Since the word list for the 1970's and the Dolch list had 189 words in common, the cumulative frequencies of these words were determined in both the AHI Corpus and the Kucera-Francis Corpus. As indicated in Table 1, approximately 2,763,051 of the 5,088,721 words in the AHI Corpus were accounted for by the 189 words common to both lists. The 31 unique words in the Dolch list resulted in another 40,469 words for a cumulative frequency of 2,803,520 words. The word list for the 1970's contained 37 unique words which accounted for an additional 130,113 words. Adding this figure to the 2,763,051 resulted in a cumulative frequency of 2,893,164 words.

It was evident that the word list for the 1970's accounted for a greater proportion of words in the AHI Corpus than the Dolch list. To test whether or not this difference was statistically significant, a one-tailed test for proportions was conducted. As indicated in Table 1, differences significant beyond the .01 level existed for both unique words and the cumulative frequency of words. On the basis of frequency the word list for the 1970's accounted for significantly more words in materials commonly used by children than the Dolch list.

A similar procedure was used to compare the Dolch list and the word list for the 1970's to the Kucera-Francis Corpus. As shown in Table 2, the word list for the 1970's accounted for a greater proportion of words in the Kucera-Francis Corpus

than the Dolch list. These differences were also tested for significance using one-tailed tests for proportions. The results were significant beyond the .01 level for both unique words and the cumulative frequency of words. On the basis of frequency, the word list for the 1970's accounted for significantly more words in adult materials than the Dolch list.

Discussion, Conclusions, and Challenge to the Future

The results of this study offer evidence that the word list for the 1970's accounts for a significantly greater proportion of words than the Dolch list in written materials encountered by both children and adults. If one uses the criterion of frequency, it is clear that the word list for the 1970's is statistically superior to the Dolch list. Not surprisingly, the word list for the 1970's (see Table 3) accounted for a greater percentage of words than the Dolch list in both the AHI Corpus and the Kucera-Francis Corpus. It should be pointed out, however, that both lists accounted for over 50 per cent of the words used in materials for children and adults.

Several conclusions are justified from the results of this study. First, the word list for the 1970's is, in fact, more useful than the Dolch list if the criterion of frequency is employed. This conclusion was also supported by several informal checks of different basal readers series frequently used in today's schools. Second, the finding that the Dolch list still accounts for over 55 per cent of the words used in materials written for children in grades three through nine and for over 50 per cent of the words frequently used in so-called adult materials offers little evidence to critics who claim that the Dolch list is passé. Certainly the vast majority of the Dolch words have withstood the test of time. Finally, although the word list for the 1970's is statistically

significant to the Dolch list, one must wonder about the practical significance of the difference.

It is altogether possible that there has been too much attention to developing word lists and not enough attention to how word lists facilitate the effective teaching of reading. A child will not become an effective reader unless he develops a large sight vocabulary - and it is obvious that the Dolch list and other recently published word lists contain many of those words. On the other hand, it is clear that knowing 220 or some magic number of words is not a sufficient condition to become an effective reader.

Word count studies can be used to demonstrate that the child who knows only thirteen words will be equipped to deal with approximately 25 per cent of the words he meets in print. While this reduces the burden of unknown words for the child, it is a far cry from making him a proficient reader. Knowing a hundred words will account for 50 per cent of the running words, but, once again, this will not make the child an efficient reader. Even knowing 2,500 words still leaves the child with approximately one unknown word in every four in a natural reading situation. Clearly, word lists quickly reach a point of diminishing returns.

These general comments on word lists are intended to stimulate a careful examination of why word lists are developed. Merely because someone develops a new word list is not, in itself, a sufficient reason for using it. The same contention is also directed at "older" word lists - like the Dolch list. Perhaps it is time to stop developing new word lists and begin to seek answers to some of the following questions:

1. What role do sight words play in the acquisition of effective and efficient reading?
2. What methods for teaching sight words are supported by research?
3. What values do word lists have for the teaching of reading?
4. How are high frequency words best learned by individuals attempting to become efficient readers?

Answers to these and related questions may help recently published word lists take on a new meaning.

Table 1

Cumulative Frequencies of Words in the AHI Corpus for the Dolch List and the Word List for the 1970's

Word List	Words Common to Both Lists	Words Unique to Both Lists	z	Cumulative Frequency	z
Dolch List (220 words)	2,763,051	40,469		2,803,520	
Word List for 1970's (226 words)	2,763,051	130,113	216.7*	2,893,164	217.1*

*significant beyond the .01 level

Table 2

Cumulative Frequencies of Words in the Kucera-Francis Corpus for the Dolch List and the Word List for the 1970's

Word List	Words Common to Both Lists	Words Unique to Both Lists	z	Cumulative Frequency	z
Dolch List (220 words)	515,035	3,836		518,871	
Word List for 1970's (226 words)	515,035	22,878	119.54*	537,913	116.5*

*significant beyond the .01 level

Table 3

Approximate Percentage of Words in the AHI Corpus and the Kucera-Francis Corpus Accounted for by the Dolch List and the Word List for the 1970's

	Per Cent of Words	
	Dolch List	Word List for the 1970's
AHI Corpus	55.09	56.85
Kucera-Francis Corpus	51.16	53.04

References

CARROLL, John B., DAVIES, Peter, & RICHMAN, Barry. The American Heritage word frequency book. Boston: Houghton Mifflin, 1971.

DOLCH, Edward W. Problems in reading. Champaign, Illinois: The Garrard Press, 1948.

DOLCH, Edward W. A basic sight vocabulary. The Elementary School Journal, February 1936, 36, 456-460.

GUSZAK, Frank J. Diagnostic reading instruction in the elementary school. New York: Harper and Row, Publishers, 1972.

HARRIS, Albert J., & JACOBSON, Milton D. Basic elementary reading vocabularies. New York: The Macmillan Company, 1972.

HILLERICH, Robert L. Word lists--getting it all together. The Reading Teacher, January 1974, 27, 353-360.

JOHNS, Jerry L. Should the Dolch list be retired, replaced, or revised? The Elementary School Journal, March 1974, 74, 375-380.

JOHNS, Jerry L. The Dolch basic word list--then and now. Journal of Reading Behavior, Fall 1971, 3, 35-40.

JOHNS, Jerry L. Updating the Dolch basic sight vocabulary for the schools of the 1970's. Paper presented at the conference of the International Reading Association, New Orleans, Louisiana, May 1-4, 1974.

JOHNSON, Dale D. A basic vocabulary for beginning reading. The Elementary School Journal, October 1971, 72, 29-34 (a).

JOHNSON, Dale D. The Dolch list reexamined. The Reading Teacher, February 1971, 24, 449-457 (b).

KUCERA, Henry, & FRANCIS, W. Nelson. Computational analysis of present-day American English. Providence, R. I.: Brown University Press, 1967.

OTTO, Wayne, & CHESTER, Robert. Sight words for beginning readers. The Journal of Educational Research, July-August 1972, 65, 435-443.

ZINTZ, Miles V. Corrective reading (2d Ed.). Dubuque, Iowa: Wm. C. Brown Company Publishers, 1972.